AD-A049 422 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
DYNKIN'S IDENTITY APPLIED TO BAYES SEQUENTIAL ESTIMATION OF A P--ETC(U)
OCT 77 C P SHAPIRO, R L WARDROP DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1795 NL

| OF |
AD
A049422

END
DATE
FILMED
3 — 78
DDC

MRC Technical Summary Report #1795

DYNKIN'S IDENTITY APPIED TO BAYES' SEQUENTIAL ESTIMATION OF A POISSON PROCESS

C. P. Shapiro and
Robert L. Wardrop

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

October 1977

(Received July 8, 1977)

See back page for 1473

Sponsored by

U.S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

DYNKIN'S IDENTITY APPLIED TO BAYES' SEQUENTIAL ESTIMATION OF A POISSON PROCESS

C. P. Shapiro and Robert L. Wardrop

Technical Summary Report #1795
October 1977

ABSTRACT

Conditional on the value of θ, $\theta > 0$, let $X(t)$, $t \geq 0$, be a homogeneous Poisson process. Sequential estimation procedures of the form $(\sigma,\hat{\theta}(\sigma))$ are considered. To measure loss due to estimation, a family of functions, indexed by p, is used: $L(\theta,\hat{\theta}) = \theta^{-p}(\theta-\hat{\theta})^2$, and the cost of sampling involves cost per arrival and cost per unit time. The notion of "monotone case" for total cost functions of a continuous time process is defined in terms of the characteristic operator of the process at the total cost function. The Bayes' sequential procedure is then derived for those cost functions in the monotone case using extensions of Dynkin's identity for the characteristic operator. The properties of these procedures are studied as sampling costs tend to 0, and the procedures are then modified and compared with procedures which are optimal among all stopping rules terminating at arrivals.

AMS(MOS) Subject Classification - Primary 62L12, 62F15; Secondary 62L15.

Key Words - Dynkin's identity, characteristic operator, Poisson process, Bayes' sequential estimation.

Work Unit Number 4 - Probability, Statistics, and Combinatorics

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

SIGNIFICANCE AND EXPLANATION

When trying to explain or analyze events that occur randomly in time or in space, one tends first to test whether the events are governed by a Poisson distribution. The classic example concerns the number of soldiers killed per year from the kick of a mule in the Prussian Army in the early 1800's, and applications have continued to this day in very many different contexts, military and otherwise.

Usually, the mean arrival or occurrence rate, θ, of a Poisson process is unknown. This paper derives optimal procedures for estimating the value of θ. The class of procedures considered is "sequential" in the sense that neither the length of time the Poisson process is observed nor the number of arrivals observed are fixed in advance. Instead, the outcome of the process prior to any point in time may be used to decide whether to continue observing the process beyond that time.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

Dynkin's Identity Applied to Bayes' Sequential Estimation of a Poisson Process

by

C.P. Shapiro and Robert L. Wardrop[1]

Michigan State University and University of Wisconsin-Madison

1. Introduction. Conditional on the value of θ, $\theta > 0$, suppose that $X(t)$, $t \geq 0$, is a homogeneous Poisson process. Sequential estimation procedures of the form $(\sigma, \hat{\theta}(\sigma))$ will be considered where σ is a stopping time with respect to $\{\mathcal{F}_t, t \geq 0\}$, with $\mathcal{F}_t$ the sigma algebra of events generated by $\{X(s), 0 \leq s \leq t\}$, and $\hat{\theta}(\sigma)$ is an $\mathcal{F}(\sigma)$ measureable function, with $\mathcal{F}(\sigma)$ the sigma algebra of events prior to σ.

A family of functions, indexed by p, $0 \leq p \leq 3$, will measure loss due to estimation:

$$L_p(\theta, \hat{\theta}) = \theta^{-p}(\theta - \hat{\theta})^2. \tag{1.1}$$

The effect of the p is to measure loss in units of θ^{-p}. The case of $p = 0$, squared error loss, has received the usual attention. Dvoretsky, Kiefer, and Wolfowitz (1953), and Hodges and Lehman (1951) have suggested $p = 1$. Using $p = 2$ causes the decision problem to be invariant under the group of scale transformations (Ferguson, 1967). The restriction of p to [0,3] is for mathematical

[1] This research was sponsored, in part, by the United States Army under Contract No. DAAG29-75-C-0024.

tractibility, and the inclusion of noninteger p values increases the computational difficulty only slightly.

The cost of sampling will involve two components: c_A, the cost of observing one arrival, and c_T, the cost of observing the process for one unit of time.

In section 2, the notion of "monotone case" for continuous time problems is defined in terms of the characteristic operator of $X(t)$ at the total cost function. Then with a gamma prior distribution on θ, the Bayes' sequential procedure is derived using extensions of Dynkin's identity when the total cost is in the monotone case. As will be seen in the section, two often used stopping times (types I and II censoring) arise as Bayes' procedures with the appropriate choice of p and the costs.

In Section 3, the large sample properties of the Bayes' sequential procedures are studied when only one cost is positive.

In Section 4, the class of rules to be considered is restricted to those stopping at an arrival, and the additional cost incurred by using a modification of the continuous time Bayes' sequential procedure rather than the optimal rule in this restricted class is seen to be no larger than $p - 2$ for $p = 0,1,2$.

2. <u>The Bayes' sequential procedures</u>. Henceforth, assume θ has prior distribution $\lambda_p(\theta) = \Gamma(\alpha_o)^{-1}\beta_o^{\alpha_o}\theta^{\alpha_o-1}e^{-\beta_o\theta}$, where $\beta_o > 0$, $\alpha_o > p$, and denote this distribution gamma (α_o,β_o). With this prior, the posterior distribution of θ given $\mathcal{F}(t)$ is gamma (α_t,β_t), where $\alpha_t = \alpha_o + X(t)$ and $\beta_t = \beta_o + t$. Using loss L_p defined in (1.1), the posterior expected loss is

$$E[L_p(\theta,\hat{\theta})|\mathcal{F}(t)] = \beta_t^{p-2}\,\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1}, \tag{2.1}$$

and the Bayes' estimator of θ given $\mathcal{F}(t)$ is

$$\hat{\theta}_p(t) = (\alpha_t-p)\beta_t^{-1}. \tag{2.2}$$

The total cost of observing the process for t units of time is

$$\mathcal{C}_p(t,X(t)) = \beta_t^{p-2}\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1} + c_A X(t) + c_T t. \tag{2.3}$$

The Bayes' sequential procedure (BSP) minimizes $E\,\mathcal{C}_p(t,X(t))$ over all pairs $(\sigma,\hat{\theta}_p(\sigma))$ where σ is a stopping time with respect to $\{\mathcal{F}_t,\ t \geq 0\}$ and $\hat{\theta}_p(\sigma)$ is an estimator of θ, and $\mathcal{F}(\sigma)$ measurable. Since the Poisson process is strong Markov, only estimators which are Bayes' given $\mathcal{F}(\sigma)$ need be considered (namely, (2.2) with t replaced by σ).

Candidates for the BSP will be infinitesimal look-ahead rules derived from the characteristic operator of $X(t)$ at $\mathcal{C}_p(t,x)$. Such rules can be shown to be Bayes' if the characteristic operator changes sign only once and if the expected cost of using procedure $(\sigma,\hat{\theta}_p(\sigma))$ can be expressed in terms of the operator for a large class of stopping times σ.

Let $f(t,x)$ be measurable and finite on $[0,\infty)^2$. Define

$$D_f(t,x,h) = \frac{E[f(t+h,X(t+h))|(t,X(t)) = (t,x)] - f(t,x)}{h} \tag{2.4}$$

and consider only those functions f such that $D_f(t,x,h)$ is uniformly bounded in $h > 0$ and the limit of D_f exists as h decreases to zero. For such f define the characteristic operator at f as

$$Af(t,x) = \lim_{h \downarrow 0} D_f(t,x,h). \tag{2.5}$$

Of particular interest is the characteristic operator at $\mathscr{C}_p(t,x)$ which is given in the lemma below.

Lemma 2.1. $A\,\mathscr{C}_p(t,x) = -\beta_t^{p-3}\Gamma(\alpha_0+x-p+1)\Gamma(\alpha_0+x)^{-1} + c_A(\alpha_0+x)\beta_t^{-1} + c_T$.

Proof. See appendix.

Say that cost $\mathscr{C}_p$ is in the monotone case if $A\,\mathscr{C}_p(t,X(t))$ changes sign at most once. Before stating when $\mathscr{C}_p$ is in the monotone case, a result about the ratio of gamma functions is needed.

Lemma 2.2. Suppose $w > 0$. Then $\Gamma(w-b)\Gamma(w)^{-1}$ is increasing in w for $b < 0$ and decreasing in w for $0 < b < w$.

Proof. Take $b > 0$, and note that $\Gamma(w-b)\Gamma(w)^{-1}$ is decreasing in w if and only if $\log\Gamma(w)$ is convex. Thus, it suffices to show that the $\log\Gamma(w)$ is convex.

Fix $0 < a < 1$, and w_1, w_2. Then $[a \log\Gamma(w_1) + (1-a)\log\Gamma(w_2)] \geq \log\Gamma(aw_1 + (1-a)w_2)$ if and only if $\Gamma(w_1)^a\Gamma(w_2)^{1-a} \geq \Gamma(aw_1 + (1-a)w_2)$. But $\Gamma(aw_1 + (1-a)w_2)$

$$= \int x^{aw_1 + (1-a)w_2 - 1} e^{-x}\, dx$$

$$= \int x^{aw_1 - a} x^{(1-a)w_2 - (1-a)} d\mu, \qquad d\mu = e^{-x}dx,$$

$$\leq ||x^{aw_1 - a}||_{1/a} ||x^{(1-a)w_2 + (1-a)}||_{1/(1-a)}, \text{ by Hölder's inequality.}$$

This last expression is equal to $\Gamma(w_1)^a\, \Gamma(w_2)^{1-a}$.

Lemma 2.3. In each of the cases (i) - (iii) below, $\mathcal{C}_p$ is in the monotone case:

(i) $c_T = 0$, $c_A > 0$, $0 \leq p < 1$;

(ii) $1 \leq p \leq 2$;

(iii) $c_A = 0$, $c_T > 0$, $2 < p \leq 3$.

Proof. For (i), $A\,\mathcal{C}_p \geq 0$ if and only if $c_A\Gamma(\alpha_t+1)\Gamma(\alpha_t-p+1)^{-1} \geq \beta_t^{p-2}$. Since the rhs is decreasing in t and the lhs is nondecreasing in t, (i) is in the monotone case. For (ii), $A\,\mathcal{C}_p \geq 0$ iff $c_A\alpha_t + c_T\beta_t \geq \beta_t^{p-2}\Gamma(\alpha_t-(p-1))\Gamma(\alpha_t)^{-1}$, where the rhs is decreasing in t and the lhs is increasing in t. For (iii), $A\,\mathcal{C}_p \geq 0$ iff $c_T\beta_t^{3-p} \geq \Gamma(\alpha_t-p+1)\Gamma(\alpha_t)^{-1}$, with the rhs nonincreasing in t and the lhs increasing in t.

From now on, only those $\mathcal{C}_p$ in the monotone case will be considered. For those p, c_A, and c_T such that $\mathcal{C}_p$ is in the monotone case, define the stopping time

$$\tau_p = \text{first } t \geq 0 \text{ such that } \quad A\,\mathcal{C}_p(t,X(t)) \geq 0 \tag{2.6}$$

Another property of ratios of gamma functions is needed to derive bounds on τ_p and $X(\tau_p)$.

Lemma 2.4.

If $w > p \geq 0$, then $\Gamma(w-p)\Gamma(w)^{-1} \geq w^{-p}$.

If $0 \leq p < 1$ and $w > p + 1$, then $\Gamma(w-p)\Gamma(w)^{-1} \leq (w-(p+1))^{-p}$.

If $p \geq 1$ and $w > p$, then $\Gamma(w-p)\Gamma(w)^{-1} \leq (w-p)^{-p}$.

Proof. Let a be in $[0,\infty)$ and define $h(w,a) = (w-a)^p\, \Gamma(w-p)\Gamma(w)^{-1}$, for $w > a$. Then $h(w,a) \leq h(w+1,a)$ if and only if

$$\log w + p \log(w-a) \leq p \log (w+1-a) + \log(w-p). \tag{2.7}$$

If $p \geq 1$, then using the concavity of logs, (2.7) holds with $a = p$; thus, $h(w,p) \leq h(w+1,p)$. If $p < 1$, then (2.7) holds with $a = p + 1$, and hence $h(w,p+1) \leq h(w+1,p+1)$. Noting that $h(w,a) \to 1$ as $w \to \infty$ for each a yields the upper bounds in the lemma. Choosing $a = 0$ reverses the inequality in (2.7) and gives the lower bound, with only $w > p$ required.

Lemma 2.5. For the cases of Lemma 2.3, the following bounds on τ_p and $X(\tau_p)$ hold:

(i) $0 \leq p < 2$, $\tau_p \leq c_A^{1/(p-2)}$,

(ii) $1 \leq p < 3$, $\tau_p \leq c_T^{1/(p-3)}$,

(iii) $0 < p \leq 2$, $X(\tau_p) \leq c_A^{-1/p}\, \beta_o^{(p-2)/p} + 1$,

(iv) $1 < p \leq 3$, $X(\tau_p) \leq c_T^{-1/(p-1)}\, \beta_o^{(p-3)/(p-1)} + 1$,

where the bounds are infinite if the cost involved is zero.

<u>Proof</u>. For (i), $A\,\mathscr{E}_p < 0$ implies that

$$\beta_t^{2-p} < c_A^{-1}\,\Gamma(\alpha_t+1-p)\Gamma(\alpha_t+1)^{-1}$$

$$\leq c_A^{-1}.$$

This implies the bound claimed. (ii) is proven in a similar fashion.

For (iii), $A\,\mathscr{C}_p < 0$ implies that

$$c_A\Gamma(\alpha_t+1) < \beta_t^{p-2}\Gamma(\alpha_t+1-p)$$

$$\leq \beta_o^{p-2}\Gamma(\alpha_t+1-p).$$

Next, if $p > 1$, then

$$c_A\beta_o^{2-p} < \Gamma(\alpha_t+1-p)\Gamma(\alpha_t+1)^{-1}$$

$$\leq (\alpha_t+1-p)^{-p} \qquad \text{by Lemma 2.4.}$$

If $0 < p \leq 1$, then $c_A\beta_o^{(2-p)} < (\alpha_t-p)^{-p}$ by Lemma 2.4. Recalling that $\alpha_t = X(t) + \alpha_o$ yields (iii). Part (iv) is proven by a similar argument.

To deduce the optimality of rules τ_p, $E\,\mathscr{C}_p(\sigma,X(\sigma))$ will be expressed in terms of $A\,\mathscr{C}_p$. Referring back to $f(t,x)$ and $Af(t,x)$ defined earlier in the section, let σ denote a stopping time with $E\sigma < \infty$. Also note that $X(t)$ is marginally a strong Markov process. If f is bounded and $f(t,X(t))$ is a.s. right continuous in t, then the following identity due to Dynkin is well known:

$$Ef(\sigma,X(\sigma)) = E\int_0^\sigma Af(t,X(t))\,dt + f(0,0). \tag{2.8}$$

For details, see Athreya and Kurtz (1973). In the present situation $\mathcal{C}_p(t,x)$ is not a bounded function, and hence some extension of (2.8) which will include $\mathcal{C}_p(t,x)$ is needed.

Lemma 2.6. Suppose f is nonnegative and continuous in (t,x). Let σ be a stopping time such that $E\int_0^\sigma |Af(t,X(t))|dt < \infty$.

(i) If $EX(\sigma) < \infty$ and if f is nonincreasing in t and nondecreasing in x such that $Ef(0,X(\sigma)) < \infty$, then (2.8) holds.

(ii) If $E\sigma < \infty$ and if f is nondecreasing in t and nonincreasing in x such that $Ef(\sigma,0) < \infty$, then (2.8) holds.

Proof. (i) Define the following sets of stopping times:

$S_1^* = \{\sigma\colon EX(\sigma) < \infty,\ Ef(0,X(\sigma)) < \infty,\ E\int_0^\sigma |Af(t,X(t))|dt < \infty\}$,

$S_2^* = \{\sigma\colon \sigma \text{ is in } S_1^* \text{ and } E\sigma < \infty\}$,

$S_3^* = \{\sigma\colon \sigma \text{ is in } S_2^* \text{ and } X(\sigma) \text{ is bounded}\}$.

First, (2.8), will be shown to hold for σ in S_3^*, then for σ in S_2^*, and finally for σ in S_1^*.

Let σ be in S_3^*. Then there exists $m < \infty$ such that $X(\sigma) < m$. Define $f_m(t,x) = f(t,x)$ on $[0,\infty) \times [0,m]$, and $= f(t,m)$ otherwise. Then (2.8) holds for f_m and σ. But $X(\sigma) < m$ implies that f_m can be replaced by f in the expression of (2.8). Thus, (2.8) holds for all σ in S_3^*.

Now take σ in S_2^*, and sequence $m_k \uparrow \infty$. Define stopping times $\sigma_k = \sigma$ if $X(\sigma) < m_k$ and $= t_k$ if $X(\sigma) \geq m_k$, where t_k is the first time that $X(t) = m_k$. Then σ_k is in S_3^* and (2.8) holds for f and σ_k. However, $Ef(\sigma_k,X(\sigma_k)) = Ef(\sigma,X(\sigma))\,[X(\sigma) < m_k] + Ef(t_k,X(t_k))\,[X(\sigma) \geq m_k]$. The first

term tends to $Ef(\sigma,X(\sigma))$ and the second term tends to 0 since $f(\sigma,X(\sigma))$ and $f(t_k,X(t_k))$ are both bounded by $f(0,X(\sigma))$, which is integrable, and $EX(\sigma) < \infty$.

In like manner $E\int_0^{\sigma_k} Af(t,X(t))\,dt = E\int_0^{\sigma} Af(t,X(t))\,dt\,[X(\sigma) < m_k]$ $+ E\int_0^{t_k} Af(t,X(t))\,dt\,[X(\sigma) \geq m_k]$, where the first term tends to $E\int_0^{\sigma} Af(t,X(t))dt$ and the second term is bounded by $E\int_0^{\sigma} |Af|[X(\sigma) \geq m_k]$ which tends to 0.

Now take σ in S_1^*, and sequence $t_k \uparrow \infty$. The truncation is now done on σ. Define $\sigma_k = \min(\sigma,t_k)$. Then σ_k is in S_2^* for all k and (2.8) holds for σ_k and f. As before, $Ef(\sigma_k,X(\sigma_k)) = Ef(\sigma,X(\sigma))[\sigma \leq t_k] + Ef(t_k,X(t_k))[\sigma > t_k]$. Note that σ in S_1^* implies $EX(\sigma) < \infty$, and hence $P(\sigma < \infty) = 1$. Using this and arguments similar to those above imply that $Ef(\sigma_k,X(\sigma_k))$ tends to $Ef(\sigma,X(\sigma))$. Finally, using similar arguments, $E\int_0^{\sigma_k} Af(t,X(t))dt$ can be shown to tend to $E\int_0^{\sigma} Af(t,X(t))dt$.

(ii) The proof is parallel to that of (i) with the roles of $X(\sigma)$ and σ reversed.

The lemma below makes precise the roles of the characteristic operator and Dynkin's identity in proving the optimality of τ_p.

<u>Lemma 2.7</u>. Consider only those $\mathcal{C}_p$ in the monotone case, and define $S_p = \{\sigma$: either (2.8) holds with $f(t,x) = \mathcal{C}_p(t,x)$ or $E\,\mathcal{C}_p(\sigma,X(\sigma)) = \infty\}$. If τ_p is in S_p and $E\,\mathcal{C}_p(\tau_p,X(\tau_p)) < \infty$, then $E\,\mathcal{C}_p(\tau_p,X(\tau_p)) \leq E\,\mathcal{C}_p(\sigma,X(\sigma))$, for all σ in S_p.

<u>Proof</u>. Let σ be a stopping time such that (2.8) holds and $E\,\mathcal{C}_p(\sigma,X(\sigma))$ is finite, and recall the definition of τ_p. Then

$$E\,\mathcal{C}_p(\sigma,X(\sigma)) - E\,\mathcal{C}_p(\tau_p,X(\tau_p)) = E\int_0^\sigma A\,\mathcal{C}_p - E\int_0^{\tau_p} A\,\mathcal{C}_p$$

$$= E[\sigma \geq \tau_p]\int_{\tau_p}^{\sigma} A\,\mathcal{C}_p - E[\sigma < \tau_p]\int_{\sigma}^{\tau_p} A\,\mathcal{C}_p$$

which is nonnegative by definition of τ_p and the monotone property of $\mathcal{C}_p$.

<u>Theorem 2.1</u>. Suppose $\mathcal{C}_p$ is in the monotone case. Then τ_p is optimal in that $E\,\mathcal{C}_p(\tau_p,X(\tau_p)) \leq E\,\mathcal{C}_p(\sigma,X(\sigma))$ for all stopping times σ.

<u>Proof</u>. In each of the cases (i)-(iii) of Lemma 2.3, the class S_p defined in Lemma 2.7 will be shown to include all stopping times σ. (i) In this case, $\mathcal{C}_p(t,x)$ is decreasing in t and increasing in x. Suppose σ is a stopping time with $EX(\sigma) < \infty$ and note that if $EX(\sigma) = \infty$ then the expected cost is infinite. Next, $\mathcal{C}_p(0,x)$ is bounded above by $\beta_0^{p-2}\,(\alpha_0+x+1-p)^{1-p} + c_A x$ using the lower bound in Lemma 2.4. Thus, $EX(\sigma) < \infty$ implies that $E\,\mathcal{C}_p(0,X(\sigma)) < \infty$. In like manner, $E\int_0^\sigma |A\,\mathcal{C}_p| < \infty$, and Lemma 2.6 (i) implies that S_p contains all σ. Part (i) is completed by noting that $E\,\mathcal{C}_p(\tau_p,X(\tau_p)) < \infty$.

For case (ii), first note that $E\,\mathcal{C}_p(\tau_p,X(\tau_p)) < \infty$. The cost is $\mathcal{C}_p(t,x) = (\beta_0+t)^{p-2}\Gamma(\alpha_0+x+1-p)\Gamma(\alpha_0+x)^{-1} + c_A x + c_T t$, where the first term is decreasing in both t and x. Using techniques similar to those in Lemma 2.6, it can be shown that (2.8) holds with $f(t,x) = (\beta_0+t)^{p-2}\Gamma(\alpha_0+x+1-p)\Gamma(\alpha_0+x)^{-1}$ and σ a stopping time such that $P(\sigma < \infty) = 1$. Then Lemma 2.6 can be applied directly to the last two terms of the cost depending on which costs are nonzero to yield S_p containing all σ.

For (iii), again note that $E\,\mathcal{C}_p(\tau_p, X(\tau_p)) < \infty$, and that the cost is now increasing in t and decreasing in x. Consider σ with $E\sigma < \infty$ since the expected cost is infinite if $E\sigma = \infty$. Then $E\,\mathcal{C}_p(\sigma, 0) < \infty$ since $E\sigma < \infty$ implies that $E\sigma^{p-2} < \infty$ for $2 < p \leq 3$. Also $|A\mathcal{C}_p|$ is bounded which implies that $E\int_0^{\sigma} |A\mathcal{C}_p| < \infty$, and Lemma 2.6 (ii) can be used to imply that S_p contains all σ.

If either $c_T = p = 0$ or $c_A = 0$ and $p = 1$, then stopping rule τ_p is type I censoring. Note that the $X(\tau_p)$ is unbounded for this censoring, while τ_p is bounded. If either $c_T = 0$ and $p = 2$ or $c_A = 0$ and $p = 3$, then τ_p is type II censoring and τ_p is unbounded while $X(\tau_p)$ is bounded. For all other τ_p both $X(\tau_p)$ and τ_p are bounded which may be an important consideration in applications. Finally, the stopping times τ_p are (computationally) easy to use, especially for integer p.

3. <u>Large sample properties of τ_p</u>. In this section the Bayes' sequential procedures $(\tau_p, \hat{\theta}(\tau_p))$, $0 \leq p \leq 3$, are examined from the sampling theory perspective. The parameter θ is considered fixed but unknown and all probabilities and expectations are conditional on θ and denoted by P_θ and E_θ, respectively. Attention will be restricted to the case when exactly one of the pair (c_A, c_T) is positive and the other is zero. Of interest is the large sample behavior of the procedures.

Write $\tau(p, c_A, c_T, \alpha_0, \beta_0) = \tau_p$ to make explicit the dependence of τ_p on the various design parameters. It is easy to verify that, for $0 \leq p \leq 2$,

$$\tau(p, c, 0, \alpha_0, \beta_0) = \tau(p+1, 0, c, \alpha_0 + 1, \beta_0). \tag{3.1}$$

In the results that follow, (3.1) will be useful in simplifying proofs.

For $c > 0$, and $1 \leq p \leq 3$, define

$$t_p^* = t_p^*(\theta,c) = \theta^{(1-p)/2}\, c^{-1/2}. \tag{3.2}$$

Lemma 3.1.

(i) Let $c_A = 0$, $c_T = c > 0$. For every $\varepsilon > 0$ and $1 \leq p \leq 3$,

$$P_\theta[|\tau_p/t_p^* - 1| > \varepsilon] \leq k \exp[c^{-1/2} D(c,\varepsilon,\theta,p)],$$

for k a constant and $D(c,\varepsilon,\theta,p) \to D(\varepsilon,\theta,p) < 0$ and finite as $c \to 0$.

(ii) Let $c_T = 0$, $c_A = c > 0$. For every $\varepsilon > 0$ and $0 \leq p \leq 2$,

$$P_\theta[|\tau_p/t_{p+1}^* - 1| > \varepsilon] \leq k \exp[c^{-1/2}\, D(c,\varepsilon,\theta,p+1)],$$

for k and D given in (i).

Proof. (i) The result is easy to obtain for $p = 1$. For $p > 1$, write $\tau = \tau_p$, and $t^* = t_p^*$. Then $P_\theta[|\tau/t^* - 1| > \varepsilon]$ equals

$$P_\theta[\tau/t^* > 1 + \varepsilon] + P_\theta[\tau/t^* < 1 - \varepsilon]. \tag{3.3}$$

The first term in (3.3) equals $P_\theta[\tau > s]$ with $s = t^*b$ and $b = 1 + \varepsilon$. By Lemma 2.4, $P_\theta[\tau > s]$ does not exceed

(3.4) $$P_\theta[X(s) < (2-\alpha_o) + c^{1/(1-p)}\beta_s^{(p-3)/(p-1)}].$$

By Bernstein's inequality for $0 < u \le 1$, (3.4) is bounded above by

$$e^2 \exp[c^{-1/2}D^+(c,\varepsilon,\theta,p,u)]$$

with

$$D^+(c,\varepsilon,\theta,p,u) = ub\theta^{(3-p)/2}[(\beta_s/s)^{(p-3)/(p-1)}b^{2/(1-p)} + (e^{-u}-1)u^{-1}].$$

Note that $(\beta_s/s)^{(p-3)/(p-1)} \uparrow 1$ as $c \downarrow 0$; $b^{2/(1-p)} < 1$; and $(e^{-u} - 1)u^{-1} \to -1$ as $u \to 0$. Thus, there exists $u_o^+ > 0$ such that for all u, $0 < u \le u_o^+$ $D^+(c,\varepsilon,\theta,p,u) <$ for all c.

The second term in (3.3) equals $P_\theta(\tau < s)$ with $s = t^*b$ and $b = 1 - \varepsilon$. A similar argument to the one above shows that $P_\theta(\tau < s)$ is bounded above by $e^{\alpha_o} \exp[c^{-1/2}D^-(c,\varepsilon,\theta,p,u)]$ with

$$D^-(c,\varepsilon,\theta,p,u) = ub\theta^{(3-p)/2}[(e^u - 1)u^{-1} - (\beta_s/s)^{(p-3)/(p-1)}\, b^{2/(1-p)}].$$

Thus, there exists u_o^- such that for all u, $0 < u \le u_o^-$, D^- is negative for all $c \le c_o$ for some $c_o > 0$. Let $u_o = \min(u_o^+,u_o^-)$ and let $D(c,\varepsilon,\theta,p) = \max[D^+(c,\varepsilon,\theta,p,u_o), D^-(c,\varepsilon,\theta,p,u_o)]$.

(ii) may be proven by the above argument and the identity (3.1).

Lemma 3.2. (i) Let $c_A = 0$, $c_T = c > 0$. For $1 \le p < 3$, τ_p/t_p^* is uniformly integrable as $c \to 0$.

(ii) Let $c_T = 0$, $c_A = c > 0$. For $0 \le p < 2$, τ_p/t_{p+1}^* is uniformly integrable as $c \to 0$.

Proof. (i) Let $Y_c = |\tau_p/t_p^* - 1|$. It suffices to show that for fixed $a > 0$, $\int_{[Y_c>a]} Y_c dP_\theta$ tends to 0 as $c \to 0$. By Lemmas 2.5 and 3.1,

$$\int_{[Y_c>a]} Y_c \, dP_\theta \le \theta^{(p-1)/2}[c^{-1/2}]^{(p-1)/(3-p)} \; k^* \exp[c^{-1/2}D(c,a,\theta,p)]$$

which tends to 0 as $c \to 0$.

(ii) can be proven with the same argument.

Theorem 3.1.

(i) Let $c_A = 0$, $c_T = c > 0$. For $1 \le p \le 3$,

$\lim_{c\to 0} \tau_p/t_p^* = 1$ (in P_θ probability) and $\lim_{c\to 0} E_\theta(\tau_p/t_p^*) = 1$.

(ii) Let $c_T = 0$, $c_A = c > 0$. For $0 \le p \le 2$,

$\lim_{c\to 0} \tau_p/t_{p+1}^* = 1$ (in P_θ probability) and $\lim_{c\to 0} E_\theta(\tau_p/t_{p+1}^*) = 1$.

Proof. (i) The first limit follows immediately from Lemma 3.1. If $1 \le p < 3$, then the second limit follows from Lemmas 3.1 and 3.2. If $p = 3$, then the definition of τ_3 implies that

$$c^{-1/2} - \alpha_0 \le X(\tau_3) \le c^{-1/2} + (3-\alpha_0).$$

Thus,

$$t_3^* - \alpha_0\theta^{-1} \le \theta^{-1} X(\tau_3) \le t_3^* + \theta^{-1}(3-\alpha_0)$$

and,

$$1 - (t_3^*)^{-1}\alpha_0\theta^{-1} \leq (t_3^*)^{-1}E_\theta\tau_3 \leq 1 + (t_3^*)^{-1}\theta^{-1}(3-\alpha_0).$$

Noting that $t_3^* \to \infty$ as $c \to 0$ completes the proof. The proof of (ii) is similar.

Theorem 3.2.

(i) Let $c_A = 0$, $c_T = c > 0$. For $1 \leq p \leq 3$,

$$\lim_{c\to 0} c^{-1/2}\, E_\theta\, \mathcal{C}_p(\tau_p, X(\tau_p)) = 2\, \theta^{(1-p)/2}.$$

(ii) Let $c_T = 0$, $c_A = c > 0$. For $0 \leq p \leq 2$,

$$\lim_{c\to 0} c^{-1/2}\, E_\theta\, \mathcal{C}_p(\tau_p, X(\tau_p)) = 2\, \theta^{(2-p)/2}.$$

Proof: (i) Write $\tau = \tau_p$, $t^* = t_p^*$. From the definition of τ,

$$c \geq \beta_\tau^{p-3}\Gamma(\alpha_\tau+1-p)\Gamma(\alpha_\tau)^{-1} \tag{3.5}$$

which implies that $c^{-1/2}E_\theta\mathcal{C}_p \leq c^{1/2}E_\theta\beta_\tau + c^{1/2}E_\theta\tau$. This upper bound tends to $2\theta^{(1-p)/2}$ by Theorem 3.1. For a lower bound, let $\tau - \varepsilon$ be the time of the $X(\tau) - 1$ st arrival. Then the reverse inequality of (3.5) is satisfied at $\tau - \varepsilon$ and $\alpha_\tau - 1$ to yield

$$c < \beta_{\tau-\varepsilon}^{p-3}\Gamma(\alpha_\tau-p)\Gamma(\alpha_\tau-1)^{-1}.$$

This implies that

$$(3.6) \qquad c^{-1/2} E_\theta \mathcal{C}_p \geq E c^{1/2} \beta_\tau [1+\varepsilon(\beta_\tau-\varepsilon)^{-1}]^{p-3}[1+(1-p)(\alpha_\tau-1)^{-1}] + c^{1/2} E_\theta \tau.$$

The function inside the expectation of expression (3.6) is uniformly integrable and tends to $\theta^{(1-p)/2}$ in probability by Theorem 3.1 and the result that both τ and α_τ tend to ∞ as c tends to 0.

(ii) First note that $c^{1/2} E_\theta X(\tau) = c^{1/2} \theta E_\theta \tau$, (by Doob's optional stopping theorem), which tends to $\theta^{(2-p)/2}$ by Theorem 3.1. Now, from the definition of τ,

$$c\alpha_\tau \geq \beta_\tau^{p-2} \Gamma(\alpha_\tau+1-p)\Gamma(\alpha_\tau)^{-1}$$

which implies that

$$c^{-1/2} E_\theta \mathcal{C}_p \leq c^{1/2} E_\theta \alpha_\tau + c^{1/2} E_\theta X(\tau).$$

This upper bound tends to $2\theta^{(2-p)/2}$. Using the same techniques as in (i) yields

$$c^{-1/2} E_\theta \mathcal{C}_p \geq E_\theta c^{1/2}(\alpha_\tau - p)[\beta_\tau(\beta_\tau-\varepsilon)^{-1}]^{p-2} + c^{1/2} E_\theta X(\tau)$$

which tends to $2\theta^{(2-p)/2}$.

Once the limiting form of τ is derived, the asymptotic distribution of $\hat{\theta}(\tau)$ can be found by straightforward methods.

Theorem 3.3. If either $c_A = 0$, $c_T = c > 0$ and $1 \leq p \leq 3$, or $c_T = 0$, $c_A = c > 0$, and $0 \leq p \leq 2$, then

$$\tau_p^{1/2}(\hat{\theta}(\tau_p) - \theta)\theta^{-1/2} \to Z \text{ (in distribution)},$$

as $c \to 0$, where Z is standard normal.

Proof. The result is an extension of the theorem on random sums of independent and identically distributed random variables by Renyi (1957). The proof given in Shapiro and Wardrop (1977) may be used here.

Theorem 3.3 says that for small c, $\hat{\theta}(\tau_p)$ has approximately a normal distribution with mean θ and variance θ/τ_p. If $c_T > 0$, $c_A = 0$, and $1 \leq p \leq 3$, Theorem 3.1 allows replacement of τ_p by its asymptotic equivalent t_p^* to give the variance of $\hat{\theta}(\tau_p)$ asymptotically equivalent to $c^{1/2}\,\theta^{(p+1)/2}$. Thus, p may be chosen by the experimenter to obtain the asymptotic variance of $\hat{\theta}(\tau_p)$ proportional to the desired power of θ. For $c_A > 0$, $c_T = 0$, $0 \leq p \leq 2$, the same argument shows that the asymptotic variance of $\hat{\theta}(\tau_p)$ is proportion to $\theta^{1+(p/2)}$.

4. Applications to discrete time rules. Let H denote the class of stopping times with respect to $\{\mathcal{F}_t, t \geq 0\}$. In previous sections the optimal stopping time in H under various cost functions was derived and studied. In some situations, one may prefer to find the optimal stopping time in the class $D \subset H$, where D is the collection of all rules which stop only at arrival times. More precisely, let $t_0 = 0$, and $t_i = \inf\{t: X(t) = i\}$, $i \geq 1$. Then a stopping time σ is in D if and only if $\sum_{i=0}^{\infty} P(\sigma = t_i) = 1$. In this section the problem of finding an optimal stopping time in D will be discussed and nearly optimal procedures will be presented. Throughout this section, take $c_T = 0$, $c_A > 0$ and $p = 0$, 1, or 2. This problem with the above sampling costs but different estimation

cost has been studied by Starr and Woodroofe (1972) from a non-Bayesian point of view. For p = 2, note that τ_2 is in D and hence is optimal in D because it is optimal in the larger class H. There appears to be no simple-to-use expression for the optimal stopping time in D for p = 0 or 1. However, nearly optimal rules in D for this case can be derived from the optimal rules in H in a natural way. For p = 0, 1, define $\sigma_p = t_{n(p)}$, where $n(p) = \inf\{m:\ t_m \geq \tau_p\}$. It is easy to see that

$$\mathcal{C}_{\sigma_0} \leq \mathcal{C}_{\tau_0} + \beta_{\tau_0}^{-2} + c_A \leq \mathcal{C}_{\tau_0} + 2c_A,$$

and

$$\mathcal{C}_{\sigma_1} \leq \mathcal{C}_{\tau_1} + c_A.$$

Thus,
$$E(\mathcal{C}_{\sigma_p}) - \inf_{\sigma\in D} E(\mathcal{C}_{\sigma}) \leq E(\mathcal{C}_{\sigma_p}) - E(\mathcal{C}_{\tau_p})$$
$$\leq c_A(2-p) \quad \text{for } p = 0, 1.$$

Thus, the expected cost incurred by using σ_p instead of the unknown optimal rule in D is at most equal to the cost of 2 - p additional observations.

Appendix

Proof of Lemma 2.1. Write

$\mathcal{C}_p(t,x) = H_p(t,x) + c_A x + c_T t$, where $H_p(t,x) = \beta_t^{p-2}\Gamma(\alpha_o+x+1-p)\Gamma(\alpha_o+x)^{-1}$, the cost due to estimation. Then, $A\mathcal{C}_p(t,x) = AH_p(t,x) + c_A(\alpha_o+x)\beta_t^{-1} + c_T$, with the last two terms being easily computed since $Ax = (\alpha_o+x)\beta_t^{-1}$. To derive $AH_p(t,x)$, write

$$E(H_p(t+h,X(t+h))-H_p(t,x)|(t,X(t)) = (t,x),\theta)$$

$$= E(\beta_{t+h}^{p-2}\Gamma(\alpha_o+x+Y+1-p)\Gamma(\alpha_o+x+Y)^{-1}-H_p(t,x)|(t,X(t)) = (t,x),\theta)$$

where Y is Poisson (θh) given θ. Express this expectation as

$r_{p,o}(\theta,h,t) + r_{p,1}(\theta,h,t) + R_{p,2}(\theta,h,t)$, with

$$r_{p,o}(\theta,h,t) = (\beta_{t+h}^{p-2} - \beta_t^{p-2})\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1}e^{-\theta h},$$

$$r_{p,1}(\theta,h,t) = [\beta_{t+h}^{p-2}(\alpha_t+2-p)\Gamma(\alpha_t+1)^{-1} - \beta_t^{p-2}\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1}]\theta h e^{-\theta h},$$

$$R_{p,2}(\theta,h,t) = \sum_{K=2}^{\infty} [\beta_{t+h}^{p-2}\Gamma(\alpha_t+K+1-p)\Gamma(\alpha_t+K)^{-1} - \beta_t^{p-2}\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1}]\frac{(\theta h)^K}{K!}e^{-\theta h}.$$

$E[R_{p,2}(\theta,h,t)|(t,X(t)) = (t,x)]$ will be shown to be o(h). Define

$$d_K(t) = \prod_{i=1}^{K} (\alpha_t - p + i)(\alpha_t - 1 + i)^{-1},$$

and note that $d_K(t) \le 1$ for $p \ge 1$, and that $d_K(t) \le (\alpha_t - p + 1)^K \alpha_t^{-K}$ for $0 \le p < 1$. In terms of $d_K(t)$,

$$R_{p,2}(\theta,h,t) = \sum_{K=2}^{\infty} (\beta_{t+h}^{p-2} - \beta_t^{p-2})\Gamma(\alpha_t - p + 1)\Gamma(\alpha_t)^{-1} d_K(t) \frac{(\theta h)^K}{K!} e^{-\theta h}.$$

Thus, for $p \ge 1$,

$$|R_{p,2}(\theta,h,t)| \le \Gamma(\alpha_t - p + 1)\Gamma(\alpha_t)^{-1} |\beta_{t+h}^{p-2} - \beta_t^{p-2}| e^{-\theta h} |e^{\theta h} - 1 - \theta h|$$

which is $o(h)$ as $h \downarrow 0$, and is also dominated by an integrable function of θ for h in a neighborhood of zero. Thus, $E(R_{p,2}(\theta,h,t) | (t,X(t)) = (t,x)) = o(h)$. Likewise, for $0 \le p < 1$,

$$|R_{p,2}(\theta,h,t)| \le \Gamma(\alpha_t - p + 1)\Gamma(\alpha_t)^{-1} |\beta_{t+h}^{p-2} - \beta_t^{p-2}| e^{-\theta h} [|e^{\theta h} - 1 - \theta h|$$

$$+ |e^{\theta h(\alpha_t - p + 1)\alpha_t^{-1}} - 1 - \theta h(\alpha_t - p + 1)\alpha_t^{-1}|]$$

which yields $E(R_{p,2}(\theta,h,t) | (t,X(t)) = (t,x)) = o(h)$.
Finally, as $h \downarrow 0$,

$$\frac{r_{p,o}(\theta,h,t)}{h} \to (p-2)\beta_t^{p-3}\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1},$$

$$\frac{r_{p,1}(\theta,h,t)}{h} \to \beta_t^{p-2}(\Gamma(\alpha_t+2-p)\Gamma(\alpha_t+1)^{-1}-\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1})\theta.$$

Using $E[\theta|(t,X(t)) = (t,x)] = (\alpha_o+x)\beta_t^{-1}$, and noting that the limit and expectation can be interchanged yields

$$h^{-1}E(r_{p,o}+r_{p,1}|(t,X(t)) = (t,x))$$

$$\to -\beta_t^{p-3}\Gamma(\alpha_t+1-p)\Gamma(\alpha_t)^{-1} \quad \text{as } h \downarrow 0$$

completing the proof.

REFERENCES

[1] Athreya, K.B. and Kurtz, T.G. (1973). A generalization of Dynkin's identity and some applications. Ann. Probability 4 570-579.

[2] Chow, Y.S., Robbins, H., and Siegmund, D. (1971). Great Expectations: The Theory of Optimal Stopping. Houghton-Mifflin, Boston.

[3] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1953). Sequential decision problems for processes with continuous time parameter. Problems of estimation. Ann. Math. Statist. 24 403-415.

[4] Ferguson, T.S. (1967). Mathematical Statistics A Decision Theoretic Approach. Academic Press, New York.

[5] Goffman, C. and Pedrick, G. (1965). First Course in Functional Analysis. Prentice-Hall, Englewood Cliffs.

[6] Hodges, J.L. and Lehmann, E.L. (1951). Some applications of the Cramer-Rao inequality. Proc. Second Berkeley Symp. Math. Statist. Prob., University of California Press.

[7] Renyi, A. (1957). On the asymptotic distribution of the sum of a random number of independent random variables. Acta. Math. 8 193-199.

[8] Shapiro, C.P., and Wardrop R.L. (1977). The continuous time Bayes' sequential procedure for estimating the arrival rate of a Poisson process and large sample properties. Technical Report No. 1753, Mathematics Research Center, University of Wisconsin, Madison.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
1795 14 MRC-TSR-1795

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER
9 Technical summary rept.,

4. TITLE (and Subtitle)
6 DYNKIN'S IDENTITY APPLIED TO BAYES' SEQUENTIAL ESTIMATION OF A POISSON PROCESS.

5. TYPE OF REPORT & PERIOD COVERED
Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10 C. P./Shapiro [redacted] Robert L./Wardrop

8. CONTRACT OR GRANT NUMBER(s)
15 DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit Number 4 - Probability, Statistics, and Combinatorics

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE
11 Oct[redacted]77 12 27p.

13. NUMBER OF PAGES
22

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Dynkin's identity
characteristic operator
Poisson process
Bayes' sequential estimation

theta theta-bar theta sigma theta-bar

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Conditional on the value of θ, $\theta > 0$, let $X(t)$, $t > 0$, be a homogeneous Poisson process. Sequential estimation procedures of the form $(\sigma, \hat{\theta}(\sigma))$ are considered. To measure loss due to estimation, a family of functions, indexed by p, is used: $L(\theta,\hat{\theta}) = \theta^{-p}(\theta-\hat{\theta})^2$, and the cost of sampling involves cost per arrival and cost per unit time. The notion of "monotone case" for total cost functions of a continuous time process is defined in terms of the characteristic operator of the process at the total cost function. The Bayes' sequential procedure is then derived for those cost functions in the monotone case using

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE squared

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

theta to the -pth power 221 200

20. ABSTRACT (continued)

extensions of Dynkin's identity for the characteristic operator. The properties of these procedures are studied as sampling costs tend to 0, and the procedures are then modified and compared with procedures which are optimal among all stopping rules terminating at arrivals.